THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Subnational probabilistic population forecasts: the example of Australia

Dr Tom Wilson

Dr Martin Bell

Discussion Paper 2003/06

Queensland Centre for Population Research
School of Geography, Planning and Architecture
The University of Queensland

Abstract

The variability of demographic trends at the subnational scale, particularly internal and international migration, renders subnational population forecasting more difficult than at the national scale. Illustrating the uncertainty of the demographic future for subnational regions is therefore a crucial element of any set of subnational population forecasts. However, subnational forecasts are currently prepared using deterministic models which fail to properly address the issue of demographic uncertainty. The traditional high, medium and low variants approach employed by many national statistical offices poses a number of problems. Probabilistic population forecasting models have the potential to overcome many of these problems, but these models have so far been limited to national level forecasts. This paper reports a first attempt to implement a probabilistic approach subnational population forecasting model in a multiregional projection framework. The paper sets out the forecasting framework, outlines the approach adopted to formulate each of the assumptions and presents probabilistic forecasts for 2002-51 for Queensland and the rest of Australia. The forecasts show a two thirds probability that Queensland's population in 2051 will be between 6.5 and 7.5 million whilst the same range for the rest of the country is 19.7 and 22.5 million. The forecasts quantify to what extent greater uncertainty exists about the demographic future at the subnational compared to the national scale.

Acknowledgments

This work was financially supported by a collaborative research agreement with the Queensland Government Office of Economic and Statistical Research.

Authors

Dr Tom Wilson
Post-Doctoral Research Fellow, Queensland Centre for Population Research
Email: tom.wilson@uq.edu.au

Dr Martin Bell
Director, Queensland Centre for Population Research
Email: martin.bell@uq.edu.au

Paper prepared for the 2003 European Population Conference, 26-30 August, Warsaw, Poland

Queensland Centre for Population Research
School of Geography, Planning and Architecture
Chamberlain Building
The University of Queensland
St Lucia, Brisbane, Queensland 4072, Australia

http://www.geosp.uq.edu.au/qcpr

Contents

List of Figures

1. Introduction

Deterministic population forecasts frequently turn out to be rather inaccurate, sometimes within the embarrassingly short period of just one or two years of their publication. This inaccuracy can be viewed as arising from both a limited understanding of the processes of demographic change as well as an inherent randomness in these processes. The conventional method of illustrating the uncertainty of the demographic future is to produce variant population forecasts with different assumptions about the future of fertility, mortality and migration. Many national statistical offices and international organisations take combinations of the variant fertility, mortality and migration assumptions to produce high, medium and low forecast variants. But while this approach would seem a sensible way of dealing with the uncertainty issue, closer inspection reveals several major shortcomings.

First, no indication is given as to the likelihood of the low and high variants coming true (Lutz and Scherbov 1998). Are the high and low variants quite likely or very unlikely? Will the future population almost certainly be within the high-low range? The variants cannot be meaningfully interpreted. Second, the future trajectories of fertility, mortality and migration are nearly always assumed to be linear or to change smoothly over time. This simply does not match what is known about past trends. Cyclical behaviour and random fluctuations are ruled out (Lee 1999). If, as is often the case, the high and low variants of fertility, mortality or migration are slowly trended in over many years from the most recently observed value then the high-low range will open up quite slowly. The chance of actual trends exceeding that high-low range in the early years of the forecast is therefore quite high (Bryant 2003). Third, the fixed relationships between the fertility, mortality and migration assumptions in variant population forecasts give high-low ranges which will vary in their probabilistic coverage from one output variable (e.g. total population) to another (e.g. the elderly dependency ratio) (Keilman et al. 2002). A fourth and related point is that fixed combinations of fertility, life expectancy at birth and international migration preclude the many alternatives (e.g. high fertility with low international migration) that could exist in the future. Probabilistic population forecasts overcome these limitations.

Over the last decade probabilistic population forecasting methods have been progressively developed and applied to a number of countries, including Australia (Wilson and Bell 2003), Austria (Lutz and Scherbov 1998), Norway (Keilman *et al.* 2002), Finland (Alho 2002), the

Netherlands (de Beer and Alders 1999), Sweden (Cohen 1986), the United States (Lee and Tuljapurkar 1994) and for world regions (Lutz *et al.* 2001, Lutz and Scherbov 2003). To date, however, there has been no concerted attempt to apply the probabilistic approach to population forecasting at the subnational scale. At one level this is not surprising since subnational forecasts add a fourth variable, internal migration, to the three that must be considered for national forecasts. Since many probabilistic forecasts deal with international migration as net migration or as uncorrelated immigration and emigration flows this reluctance to address migration within countries is perhaps to be expected. At the same time, the imperative for a probabilistic approach is in fact more pressing since uncertainty rises as population size falls. Not only is there one extra variable to be considered it is also the most volatile of the components of population change. At the regional level, internal migration fluctuates widely in response to numerous factors, including the housing market, employment opportunities, state budget crises, and international migration. There is therefore an even greater need at the subnational scale for population forecasts to be produced with confidence intervals. If probabilistic forecasting at the subnational level simply involved the addition of one extra variable in a uniregional projection framework it could be addressed by a straightforward extension of the procedures designed for international movements. As the work of Rogers (1990) clearly demonstrates however, uniregional models notoriously fail to capture the dynamics of inter-regional migration. If internal migration is to be taken into account, a multiregional framework is called for. This adds a new dimension to the task of probabilistic forecasting.

This paper sets out one approach designed to develop methods for producing subnational probabilistic population forecasts in a biregional projection framework. The probabilistic forecasting model is described in the following section. The model was applied to Australia to generate 2002-based population forecasts for Queensland and the Rest of Australia up to 2051, and the results from this application are presented in section 3. The paper concludes by noting some challenging issues which need to be tackled in the development of subnational probabilistic population forecasting methods.

2. The forecasting model

2.1 Summary of the model

The population forecasts were produced using a probabilistic version of a two-region multiregional cohort component model (Rogers 1995, Rees 1984). The model was operationalised in a fortran 95 program which proceeds in five main stages.

Stage 1: 5,000 random simulations for summary indicators of each component of demographic change are prepared for the 2003-51 forecast period. These summary indicators are: the Total Fertility Rate (TFR), life expectancy at birth (e_0), the Gross Migraproduction Rate (GMR) as a summary measure of age-specific internal migration rates, total immigration and total emigration. The methods used to simulate each indicator are described below.

Stage 2: Age-specific detail is added to each component indicator for each forecast year using fixed age schedules.

Stage 3: The 5,000 sets of indicators are combined elementwise (fertility simulation 1 with e_0 simulation 1, etc.) and run through the cohort component model in to generate 5,000 sets of forecast populations.

Stage 4: The forecast populations for each region are assembled into database format, giving two matrices of 5,000 simulations by 49 years, by 2 sexes, and by 101 single years of age.

Stage 5: Forecast distributions are calculated for each selected measure (e.g. total population, median age, the percentage aged 65+, dependency ratios, etc.). This approach allows all-Australia forecast populations and derived measures to be calculated by adding together the output matrices for the two regions (Queensland's forecasts from the *n*th simulation are added to the Rest of Australia's forecasts from the *n*th simulation).

The probabilistic nature of the model comes from the 5,000 random trajectories of each of the summary indicators of demographic change. How these are produced and how the correlations between variables and regions are incorporated is now explained.

2.2 Generation of the forecast distributions of the summary indicators

Random trajectories for 2003-2051 are first produced for the Rest of Australia region for the TFR, female e_0 and immigration and for the Rest of Australia to Queensland GMR. These random trajectories are created by a random walk with drift model, i.e.

$$Y(t) = Y(t-1) + \varepsilon^{Y}(t) + drift^{Y}(t) \qquad (1)$$

where

Y denotes the variable of interest,

(t) is a one year period of time,

ε^{Y} is the random year on year difference for variable Y, and

$drift^{Y}$ is a user-defined amount which changes the median of variable Y's distribution.

The random walk approach has been adopted in some previous probabilistic forecasting models (e.g. de Beer and Alders 1999). As is common in other probabilistic work, the median trajectory for each summary indicator is set in advance. The values used in the forecasts for this paper are described in section 3.

Because of the large distribution of year on year immigration differences, the forecast distribution for immigration would become unrealistically wide if ceiling and floor limits were not applied. Therefore, if during any simulation immigration exceeds a specified proportion of the set trend, that simulation for the whole 2003-2051 period is rejected and another sample path generated. This continues until 5,000 trajectories which fall between the specified limits are produced. This practice follows Keilman *et al.* (2002).

To capture the close correlation between male and female e_0, male e_0 trajectories are obtained by using equation 1 to produce values of male e_0 as a proportion of female e_0. This is achieved by setting

$$Y = \frac{e_{0,m}}{e_{0,f}}. \qquad (2)$$

where

m denotes male, and

f is female.

The random proportions are then multiplied by the female e_0 values to give the 5,000 male e_0 trajectories.

Emigration in Australia is strongly correlated with immigration, and most strongly correlated with immigration two years earlier. Although a model where emigration is a random proportion of immigration two years earlier would be attractive because of its simplicity, year on year variability in emigration is smaller than for immigration, which calls for a more sophisticated model. To achieve this emigration trajectories were produced using equation 1, but with the year on year variation calculated as:

$$\varepsilon^{E}(t) = \varepsilon^{I}(t-2)\ \alpha^{E} + x^{E}(t) \tag{3}$$

where

ε^{E} denotes the year on year difference for emigration,

ε^{I} is the random year on year difference for immigration,

α^{E} is a scaling factor to change the size of the ε^{I} distribution, and

x^{E} is a random error term because the correlation between ε^{E} and ε^{I} is not perfect.

Strong correlations in the summary indicators exist between Queensland and the Rest of Australia, and these spatial correlations were incorporated as follows. The TFR trajectories for Queensland were produced as random proportions of the TFR for the rest of Australia. In equation 1

$$Y = \frac{TFR_{Qld}}{TFR_{RoA}}. \tag{4}$$

where

Qld denotes Queensland, and

RoA is the Rest of Australia region.

In a similar way Queensland's female e_0 values were modelled as random proportions of the Rest of Australia's female e_0 values, whilst the male e_0 values are modelled as random proportions of the Rest of Australia's male e_0 values. In equation 1 therefore

$$Y = \frac{e_{0,Qld,s}}{e_{0,RoA,s}}. \tag{5}$$

where

s denotes sex.

Past internal migration trends show a clear correlation between migration flows into Queensland from the Rest of Australia and those in the opposite direction. To capture this, the GMR for Queensland to the Rest of Australia was linked to the GMR for counter-flows using the random walk with drift model (equation 1) but where the year on year difference terms were defined as:

$$\varepsilon^{GMR}_{Qld-RoA}(t) = \varepsilon^{GMR}_{RoA-Qld}(t)\ \alpha^{GMR}_{Qld-RoA} + x^{GMR}_{Qld-RoA}(t) \qquad (6)$$

where

$\varepsilon^{GMR}_{Qld-RoA}$ represents the year on year difference for the Queensland to Rest of Australia GMR,

$\varepsilon^{GMR}_{RoA-Qld}$ is the random year on year difference for the Rest of Australia to Queensland GMR,

α is a scaling factor because the $\varepsilon^{GMR}_{Qld-RoA}$ and $\varepsilon^{GMR}_{RoA-Qld}$ distributions vary in size, and

$x^{GMR}_{Qld-RoA}$ is random error.

Immigration to Queensland was likewise linked to immigration to the Rest of Australia, again using equation 1, and defining the year on year immigration differences as:

$$\varepsilon^{I}_{Qld}(t) = \varepsilon^{I}_{RoA}(t)\ \alpha^{I}_{Qld} + x^{I}_{Qld}(t) \qquad (7)$$

Emigration for Queensland was linked to Queensland's immigration using equation 3.

3. Forecast assumptions

In order for the model to produce the forecast distributions of the summary indicators of demographic change, three sets of variables had to be specified:
(i) a trajectory describing the most likely future for each of the summary demographic indicators. These formed the medians of the forecast distributions.
(ii) the distributions of the year on year differences in the summary indicators.
(iii) floor and ceiling limits to prevent very unlikely or nonsensical values being included.
The parameter values adopted for each indicator are discussed in the text below.

3.1 Fertility

Both past trends and theory (McDonald 2003) suggest that Australia's TFR will continue to fall in coming years. Queensland's TFR has traditionally been very slightly higher than that of the rest of Australia. In these forecasts the most recently observed TFR values for Queensland (1.79 in 2001) and the rest of Australia (1.73) are assumed to decline gradually to 1.65 and 1.60 respectively in 25 years' time.

Taking the view that the era of below-replacement fertility with small annual fluctuations in the TFR is likely to continue, we chose standard deviations for the year on year difference term (ε^Y in equation 1) based on the distribution of year on year differences since 1975 when Australian fertility fell below replacement level. For the Rest of Australia's TFR the standard deviation of year on year differences was 0.040. Queensland's TFR has been slightly more variable (the standard deviation of the year on year differences since 1976 being 0.054) so the random error term in equation 1 was set to give a slightly wider distribution.

3.2 Mortality

Despite being one of the lowest mortality countries in the world, life expectancy at birth increases in Australia have been substantial in recent decades (ABS 2002). We take a largely optimistic view, following to some extent Oeppen and Vaupel (2002) but acknowledging that future life expectancy gains may well be more difficult to achieve (Kannisto 2001). The annual average e_0 gains of 0.22 years (females) and 0.31 years (males) over the last two decades are assumed to decline gradually to half this value by mid-century. Mortality differences between Queensland and the rest of Australia are shown by recent data be to very minor so for the purposes of these forecasts it has been assumed that no differences exist. Life expectancy at birth for both regions is therefore assumed to rise from 2001-02 values of 77.5 (males) and 82.9 (females) to 88.8 (males) and 91.2 (females) by 2050-51.

The year on year variability in e_0 is small, with the standard deviation of the Rest of Australia's female e_0 year on year variation being 0.32 years. This was assumed to continue into the future. Similar year on year differences have been observed for male e_0 and for the e_0 values for Queensland. These three e_0 values are forecast as random proportions of the Rest of Australia's female e_0 and the year on year proportion distributions were set to ensure roughly the same e_0 year on year difference distributions.

3.3 Internal migration

Past data show that rates of migration from Queensland to the rest of Australia have remained fairly stable over past decades, though rates of migration to Queensland have increased steadily. For these forecasts the GMR for migration to Queensland was therefore set to increase slowly over time from its 2001-02 base year value of 0.6 to 0.725 by 2050-51, whilst the median GMR for migration from Queensland was held constant at its base year value of 1.85. In the absence of a time series of GMRs, the random year on year GMR difference terms were set to give forecast distributions of migration flows that roughly matched past trends.

3.4 Immigration

Immigration to Australia over the past 50 years has been characterised by a long-term increase with large annual fluctuations. The increase has been driven to a large extent by non-permanent immigrants, such as students and workers on short-term contracts. In our forecasts it has been assumed Australia will continue to be attractive to non-permanent migrants, with the immigration trend continuing to rise at the 1951-2001 annual average of 3,000 per year, 2,500 in the Rest of Australia and 500 in Queensland. Because recent immigration figures are well above the long-run trends, the launch values of immigration were reduced to bring them more in line with historical figures, Queensland's 2001-02 base year immigration being set to 50,000 and the Rest of Australia's to 250,000.

The year on year variability of the Rest of Australia's immigration over the period 1971-2001 has a standard deviation of 22,100. A figure of 22,000 was used in the forecasts. Upper and lower limits of 0.35 times the median were found to give a reasonable forecast distribution. Due to the smaller immigration flows to Queensland the annual fluctuations are also smaller. A value of 0.18 was chosen for α^{I}_{Qld} (equation 7) which gave a forecast distribution of immigration of comparable width to past trends.

3.5 Emigration

Like immigration, the general trend in emigration from Australia in recent decades has shown a substantial increase, albeit with annual fluctuations and temporary reversals of trend. Given that much of the international migration increase in recent decades has been driven by non-permanent migrants it is not surprising that a close correlation exists between emigration and immigration, the strongest correlation being observed between emigration and immigration from two years earlier. The emigration figures recorded in recent years are historically high and the base year values have been reduced slightly to accord more closely with historical data. Queensland's 2001-02 base year emigration was set to 33,000 and the Rest of Australia's to 167,000.

Emigration from the Rest of Australia exhibits smaller annual fluctuations so α^{E}_{RoA} was set to 0.6. The same proportion was used for α^{E}_{Qld}.

4. Results

4.1 Total populations

Figure 1 presents the forecast distributions for Queensland, the Rest of Australia and Australia as a whole. The median forecasts for 2051 are, respectively, 7.0, 21.1 and 28.1 million, with the two thirds probability ranges covering 6.5-7.5 million in Queensland (30% of the 2051 median forecast), 19.7-22.5 million for the Rest of Australia (26% of the median) and 26.3-29.9 million for Australia (25% of the median). These forecasts compare with the 1999-based Australian Bureau of Statistics medium projections for 2051 of 6.1 million for Queensland, 19.3 million for the Rest of Australia and 25.4 million for Australia as a whole (ABS, 2000). The ABS projections are lower than our median forecasts due to conservative mortality assumptions and lower net international migration gains, and for Queensland, a lower net balance of internal migration in the ABS work.

4.2 Age-sex structures

For many users of population forecasts the age-sex structure is as important as, or more important than, total numbers. The forecast distributions of the age-sex structure for Queensland, the Rest of Australia and Australia in 2026 are illustrated in Figure 2. Greater detail on the elderly is shown in Figure 3 for the 65-84 age group and Figure 4 for the elderly 85+ age group. The graphs emphasise the fact that although considerable population ageing is certain, the extent of this ageing remains quite uncertain, especially for the very old.

4.3 Components of population change

The forecast distributions of births, deaths, internal migration, net internal migration, immigration, emigration and net international migration are displayed in Figures 5 to 11. The wide forecast intervals for births (Figure 5) reflect cumulative uncertainty over the generations. Not only is there uncertainty over the TFR, but from 2020 onwards some births are being produced by mothers whose generation size is also uncertain because they themselves were born after 2002. In contrast, the forecast distributions for deaths are narrower (Figure 6). Partly this is due to relatively narrow forecast distributions for e_0, but it is also because for the whole 2002-51 forecast horizon all those in the highest mortality ages were alive in 2002 (although not necessarily resident in the designated region). The internal migration forecast distributions (Figure 7) are relatively wide, reflecting the large variability of these flows in the past. The net internal migration distributions (Figure 8) are also fairly wide, but much narrower than they would be if no correlation existed between the Queensland to Rest of Australia and Rest of Australia to Queensland flows. Following past trends, the immigration distributions (Figure 9) are wider than the emigration distributions (Figure 10). The high correlation between immigration and emigration means that the net international migration distributions (Figure 11) are much narrower than would be the case with a looser connection between the two international migration flows.

5. Conclusion

This paper has described an initial attempt at producing a subnational probabilistic population forecasting model using a multiregional framework. Given the greater uncertainty of population futures at the subnational scale the need to replace deterministic projections with probabilistic forecasts at this scale is crucial. Although much work remains to be done, the model described here includes gross migration flows in place of net migration, and captures the important correlations that exist between migration flows and between variables across regions. The graphical output demonstrates that the method produces sensible sets of forecast intervals. The next step is to move beyond a biregional application to a model that can handle all eight Australian states and territories.

Literature cited

ABS (2002) *Deaths, 2001*. Canberra: ABS.

ABS (2000) *Population Projections, Australia: 1999 to 2101*. Canberra: ABS.

Lee, R. (1999) Probabilistic approaches to population forecasting. In Lutz, W., Vaupel, J.W. and Ahlburg, D.A. (eds) *Frontiers of Population Forecasting*. Supplement to population and Development Review volume 24. New York: Population Council; 156-190.

Lutz, W. and Scherbov, S. (1998) An expert-based framework for probabilistic national population projections: the example of Austria. *European Journal of Population* 14: 1-17.

Bryant, J. (2003) Can population projections be used for sensitivity tests on policy models? New Zealand Treasury Working Paper 03/07. Wellington, New Zealand.

Keilman, N., Pham, D.Q. and Hetland, A. (2002) Why population forecasts should be probabilistic – illustrated by the case of Norway. *Demographic Research* 6.15. http://www.demographic-research.org

Alho. J. (2002) The population of Finland in 2050 and beyond. Discussion Paper no. 826, The Research Institute of the Finnish Economy, Helsinki.

Wilson, T. and Bell, M. (2003) Australia's uncertain demographic future. Queensland Centre for Population Research Discussion Paper.

De Beer, J. and Alders, M. (1999) Probabilistic population and household forecasts for the Netherlands. Paper prepared for the joint ECE-EUROSTAT work session on demographic projections, Perugia, Italy, 3-7 May 1999.

Cohen, J.E. (1986) Population forecasts and confidence intervals for Sweden: a comparison of model-based and empirical approaches. *Demography* 23: 105-126.

Lee, R. and Tuljapurkar, S. (1994) Stochastic population forecasts for the United States: beyond high, medium and low. *Journal of the American Statistical Association* 89: 1175-1189.

Lutz, W., Sanderson, W. and Scherbov, S. (2001) The end of world population growth. *Nature* 412: 543-545.

Lutz, W. and Scherbov, S. (2003) The end of population growth in Asia. *Journal of Population Research* 20.1: 125-141.

Rees, P. (1984) Spatial population analysis using movement data and accounting methods: theory, models, the 'MOVE' program and examples. Working paper 404/Computer manual 23, School of Geography, University of Leeds.

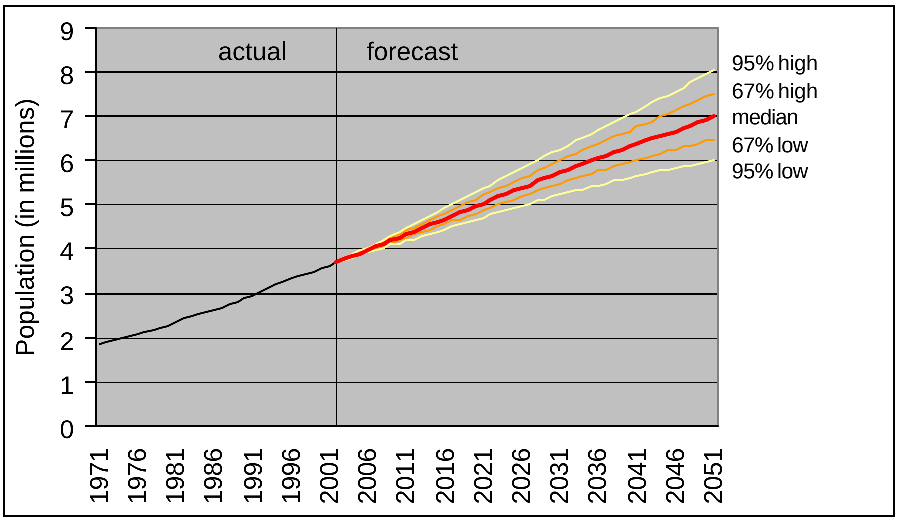
Rogers, A. (1990) Requiem for the net migrant. *Geographical Analysis* 22: 283-300.

Rogers, A. (1995) *Multiregional Demography: Principles, Methods and Extensions*. Chichester: John Wiley.

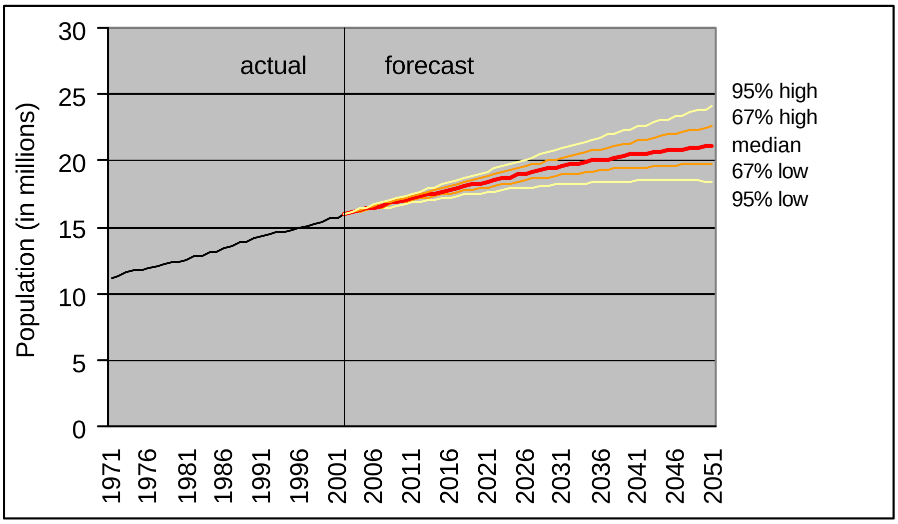
McDonald, P. (2003) Australia's future population: population policy in a low-fertility society. Chapter 11 in Khoo, S. and McDonald, P. (eds) *The Transformation of Australia's Population 1970-2030*. Sydney: University of New South Wales Press; 266-279.

Oeppen, J. and Vaupel, J.W. (2002) Broken limits to life expectancy. *Science* 296: 1029-1031.

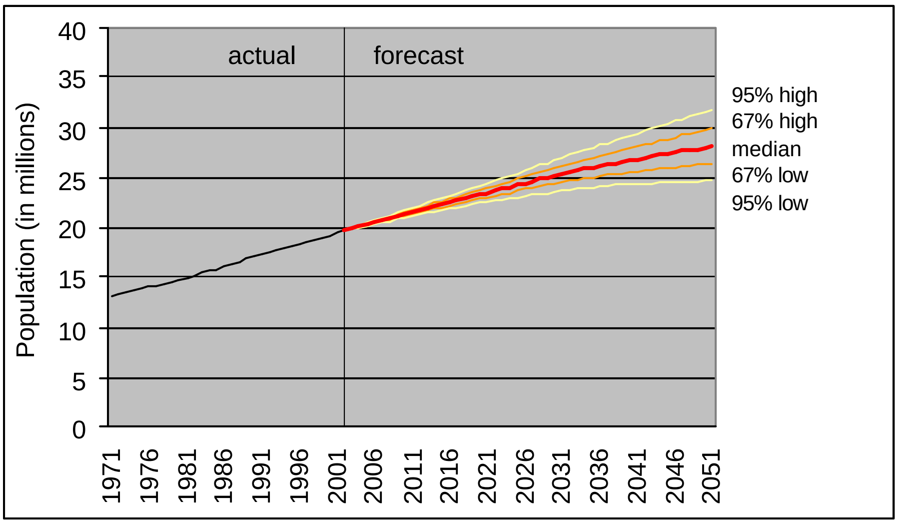
Kannisto, V. (2001) Mode and dispersion of the length of life. *Population: An English Selection* 13: 159-172.



(a) Queensland

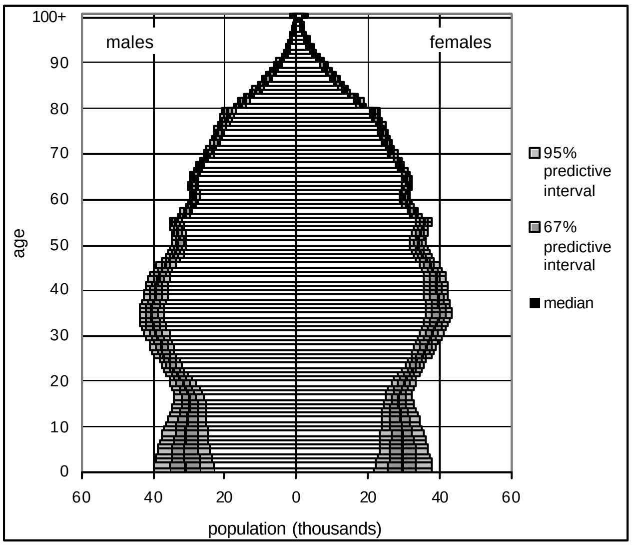


(b) Rest of Australia

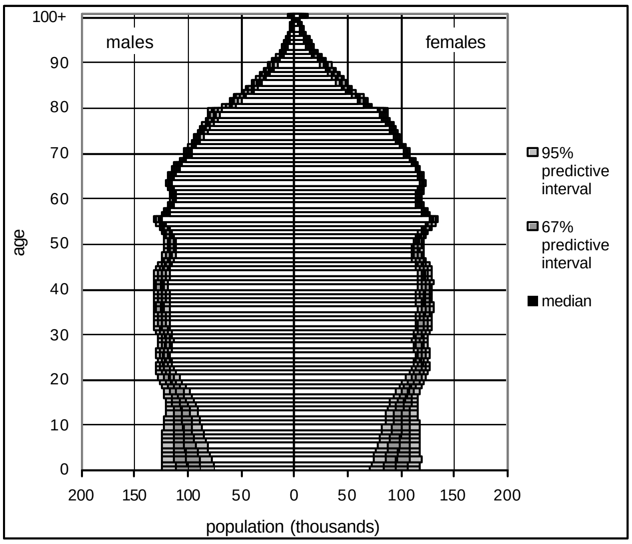


(c) Australia

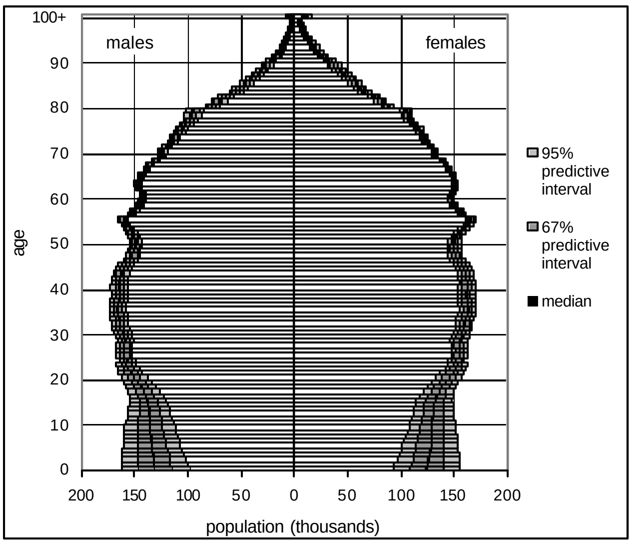
Figure 1 Observed and forecast populations, 1971-2051



(a) Queensland

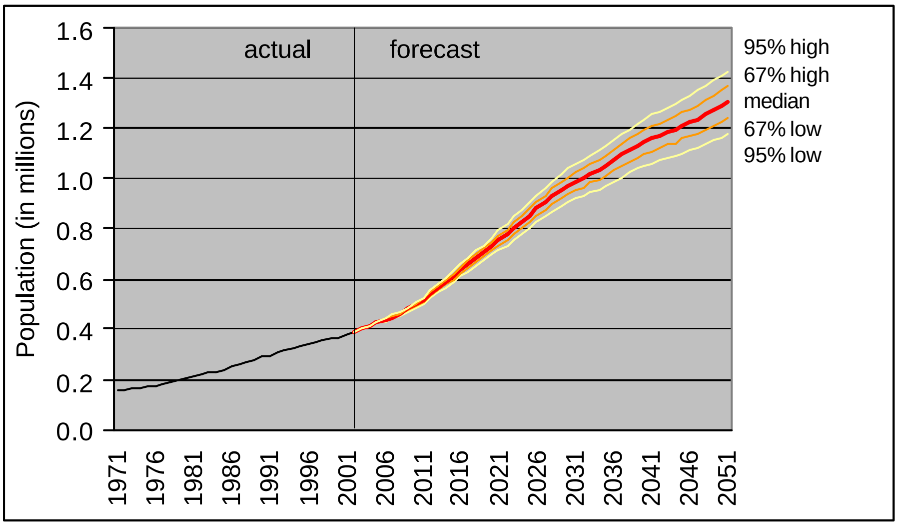


(b) Rest of Australia

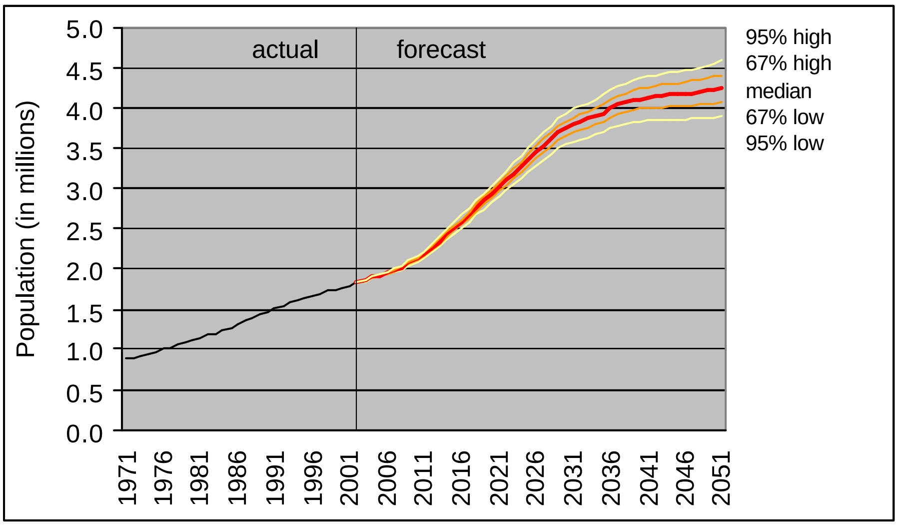


(c) Australia

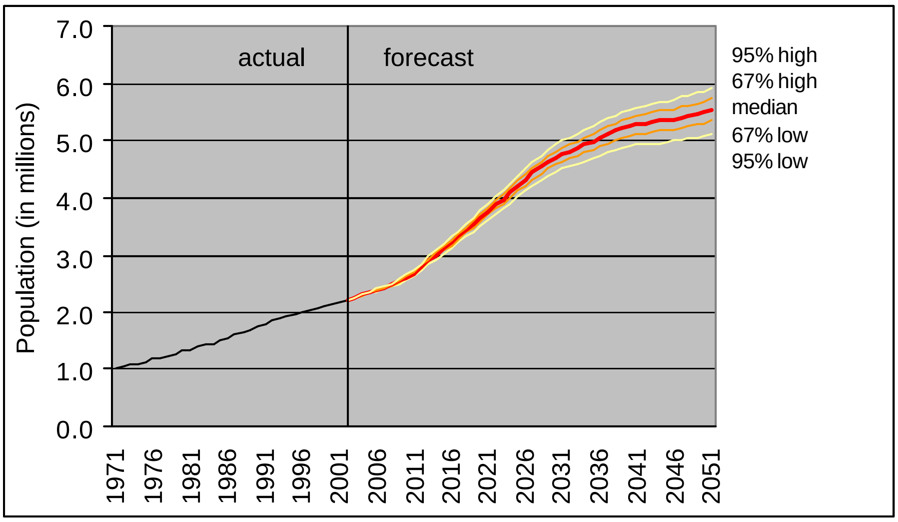
Figure 2 Forecast age-sex distributions, 2026



(a) Queensland

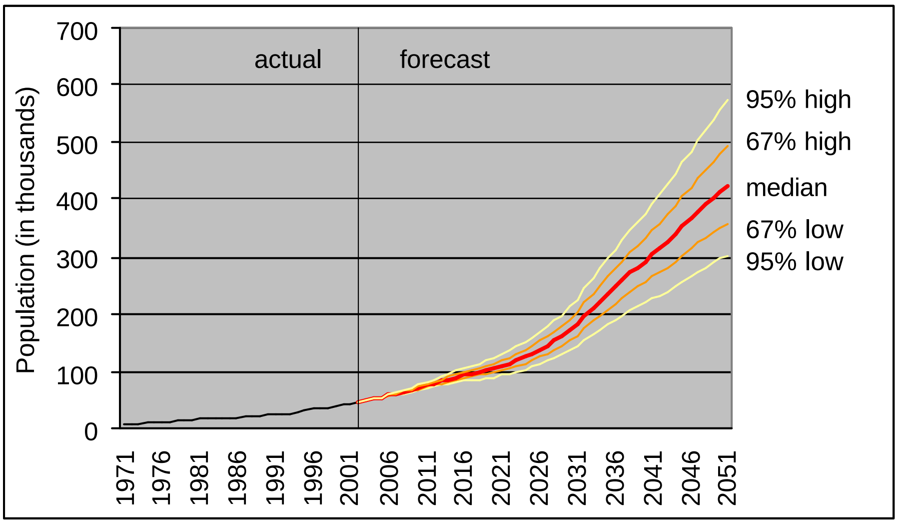


(b) Rest of Australia

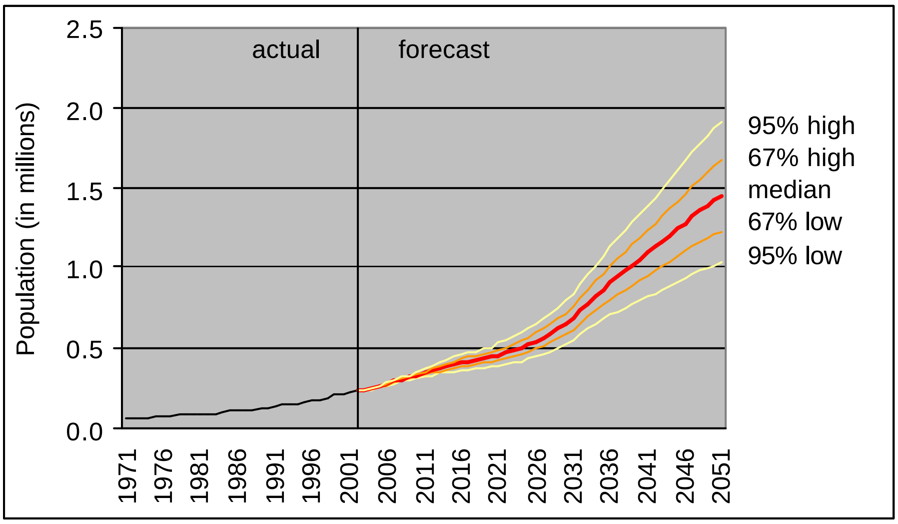


(c) Australia

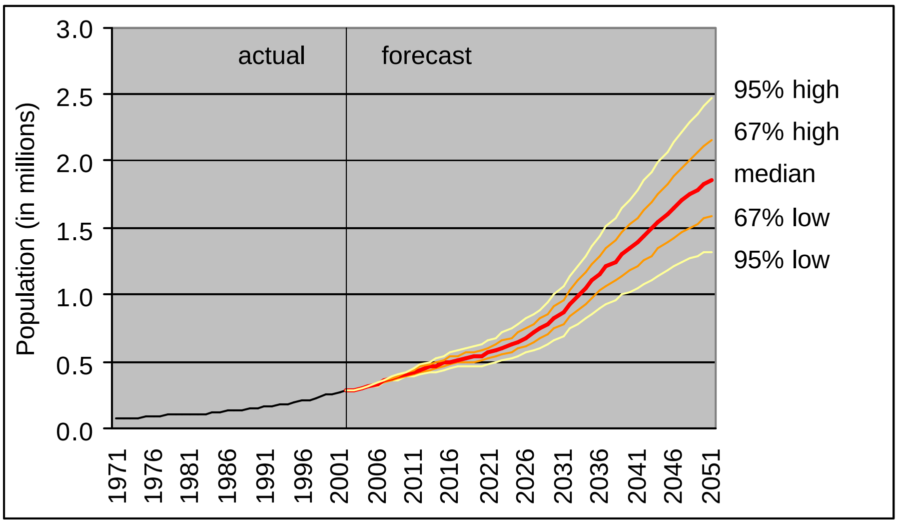
Figure 3 Observed and forecast population aged 65-84, 1971-2051



(a) Queensland

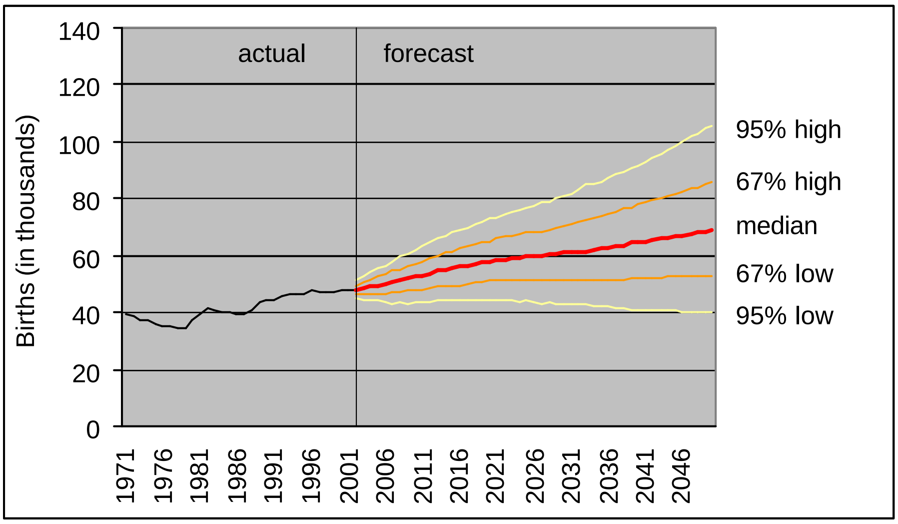


(b) Rest of Australia

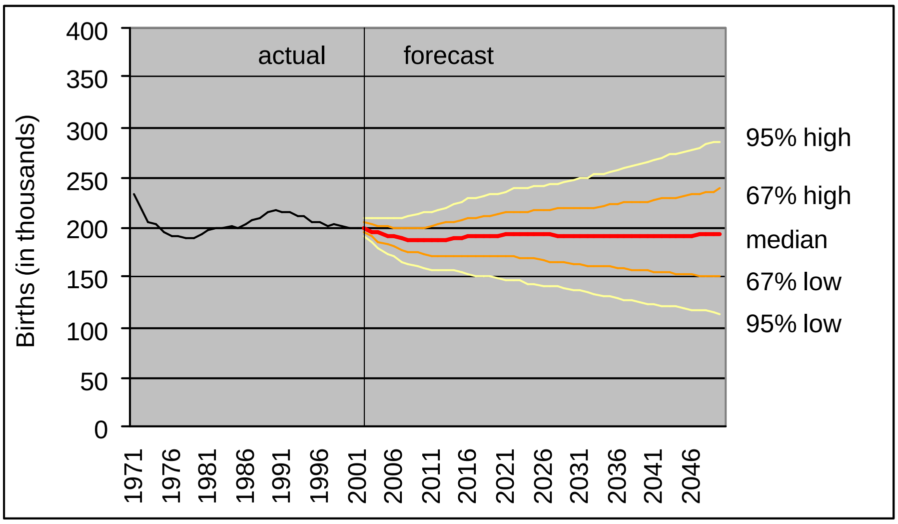


(c) Australia

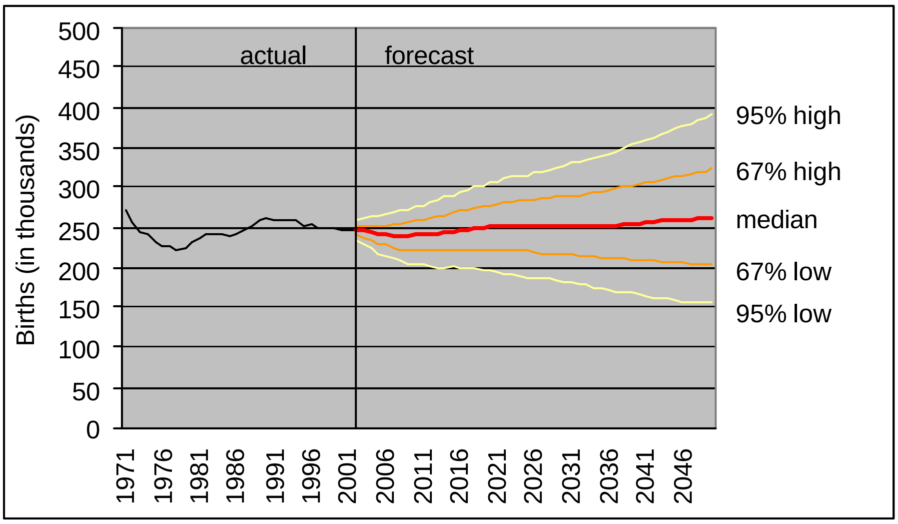
Figure 4 Observed and forecast population aged 85+, 1971-2051



(a) Queensland

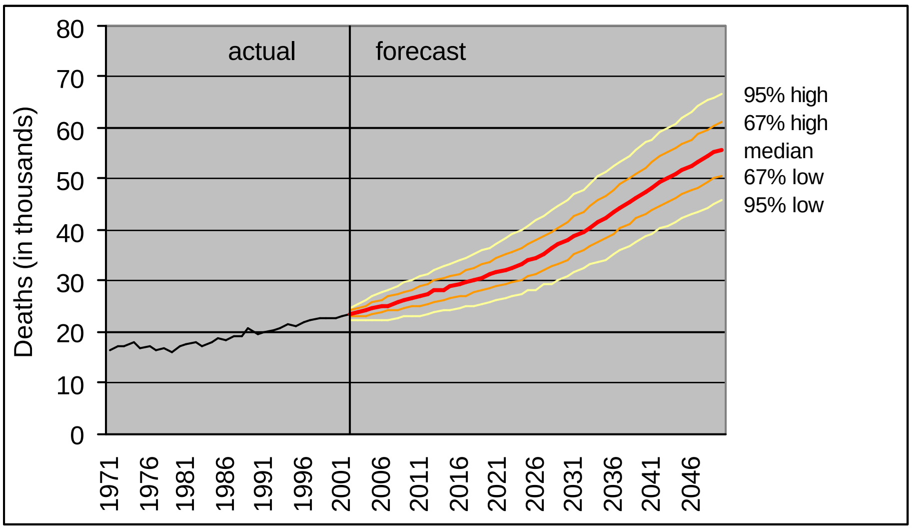


(b) Rest of Australia

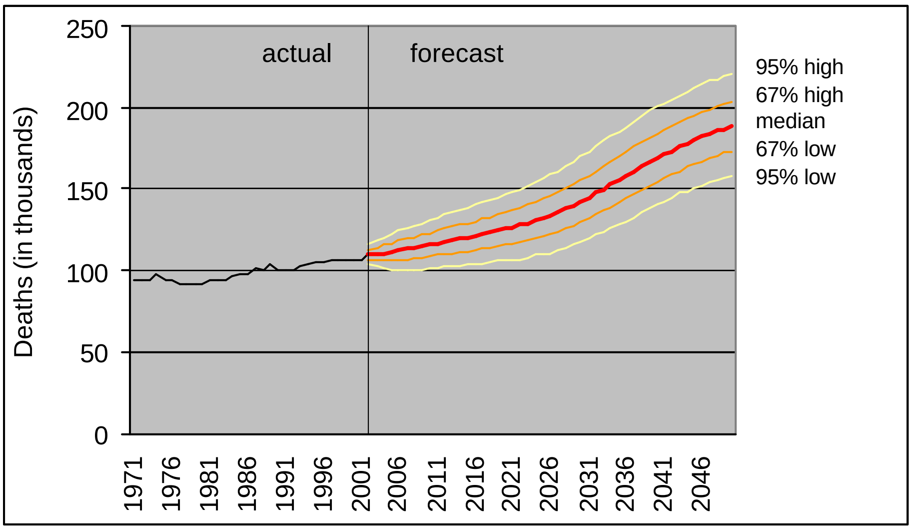


(c) Australia

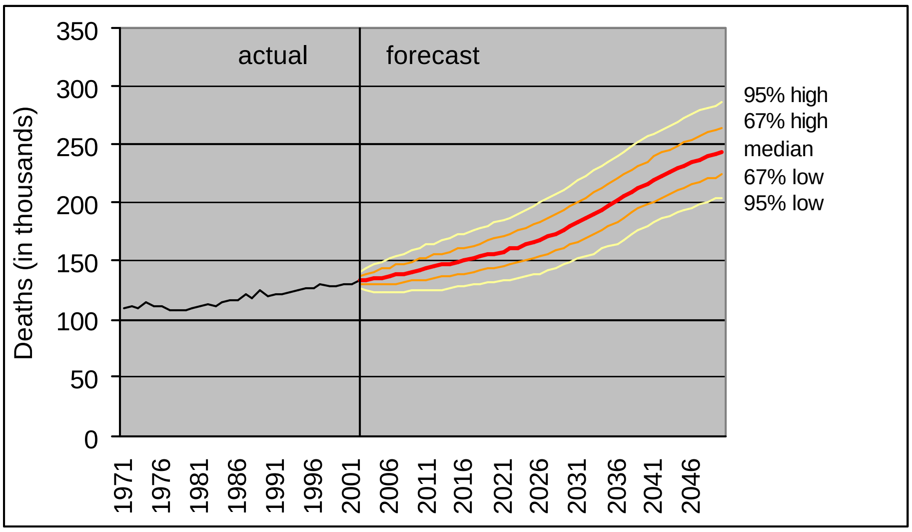
Figure 5 Observed and forecast births, 1971-2051



(a) Queensland

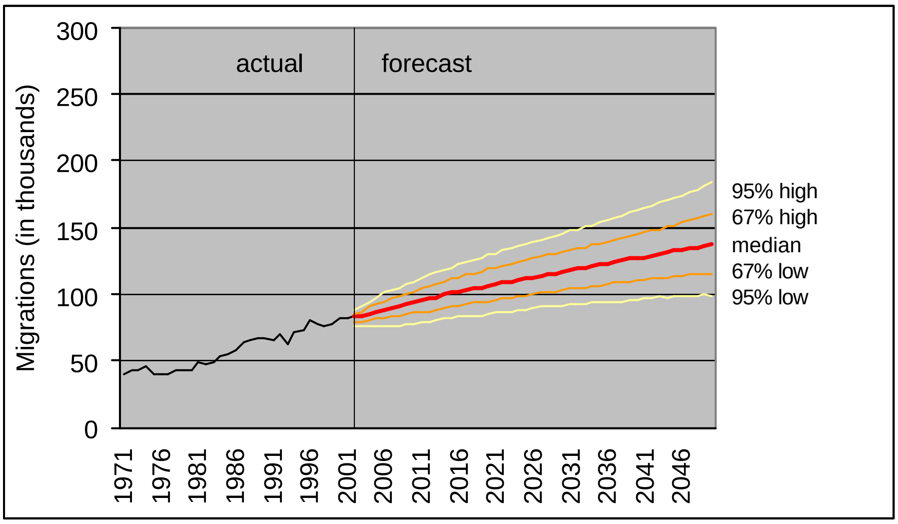


(b) Rest of Australia

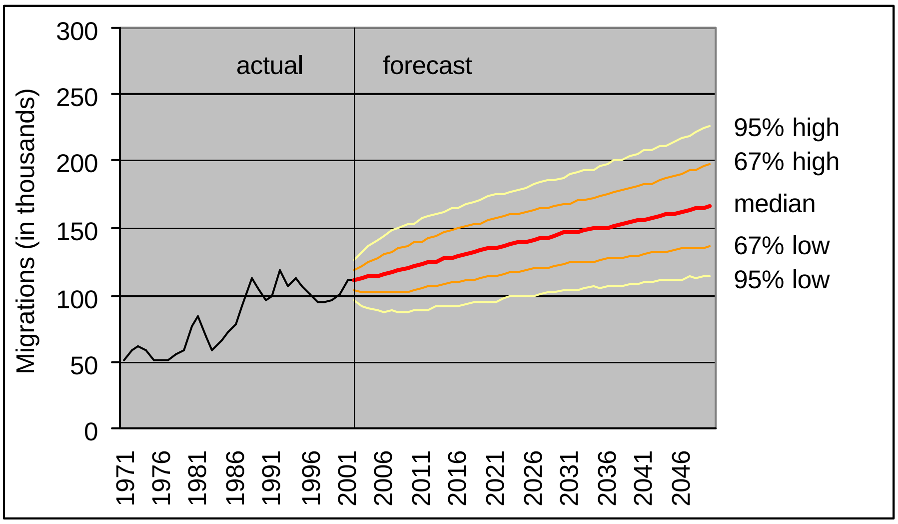


(c) Australia

Figure 6 Observed and forecast deaths, 1971-2051

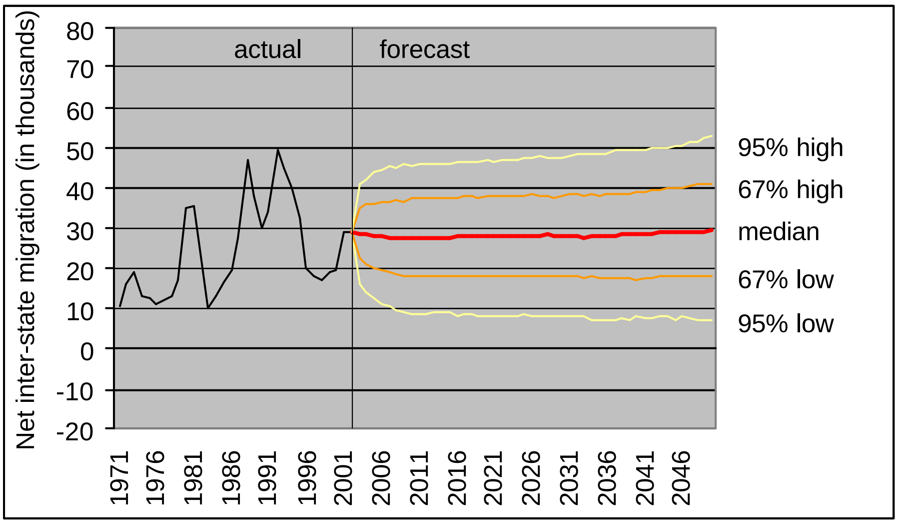


(a) Queensland to rest of Australia

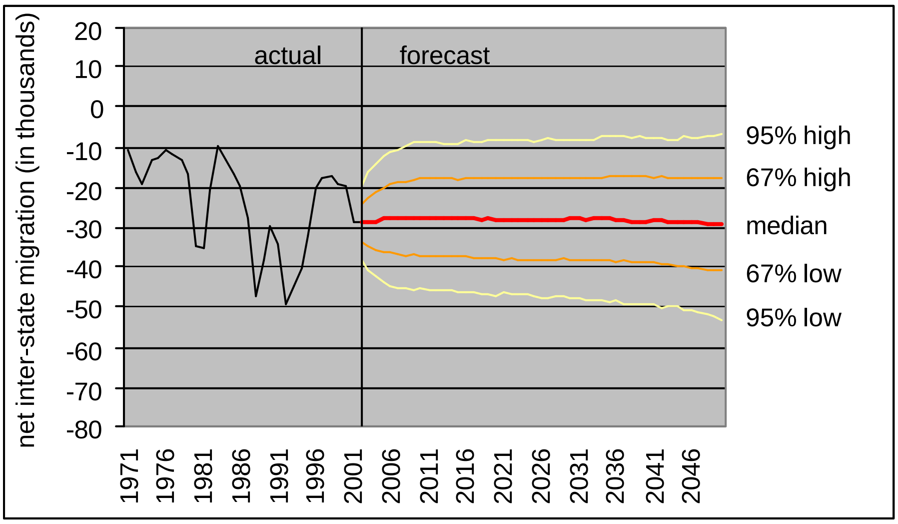


(b) Rest of Australia to Queensland

Figure 7 Observed and forecast internal migration flows, 1971-2051

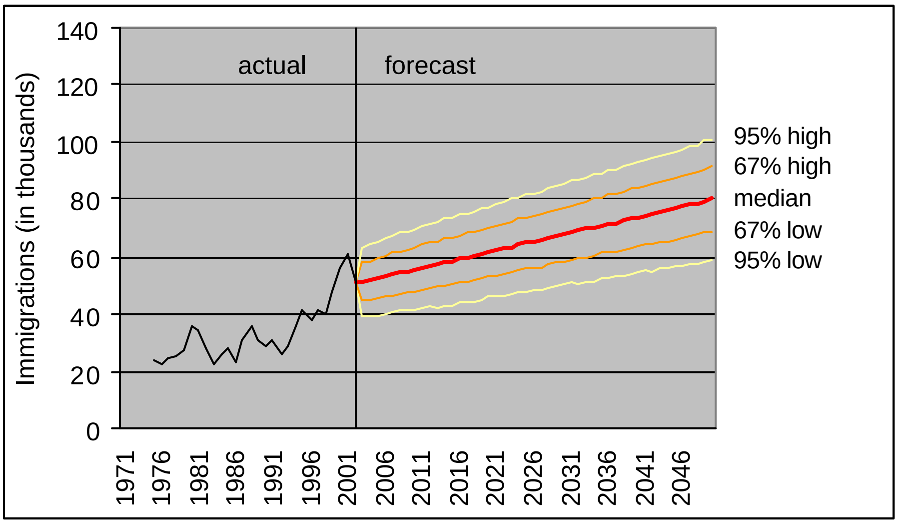


(a) Queensland

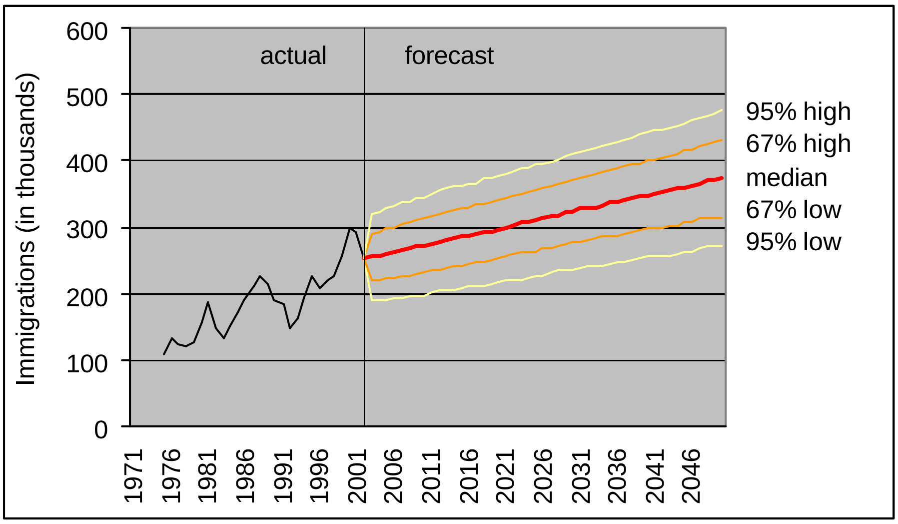


(b) Rest of Australia

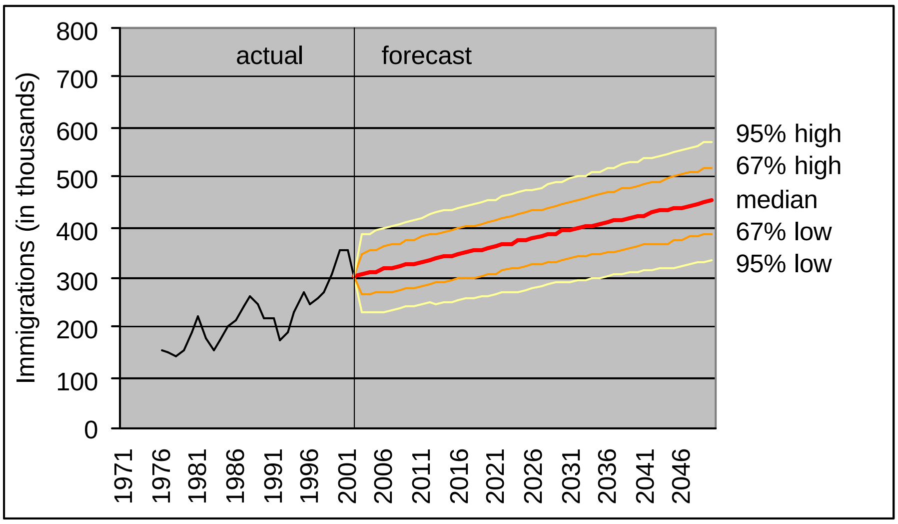
Figure 8 Observed and forecast net interstate migration, 1971-2051



(a) Queensland

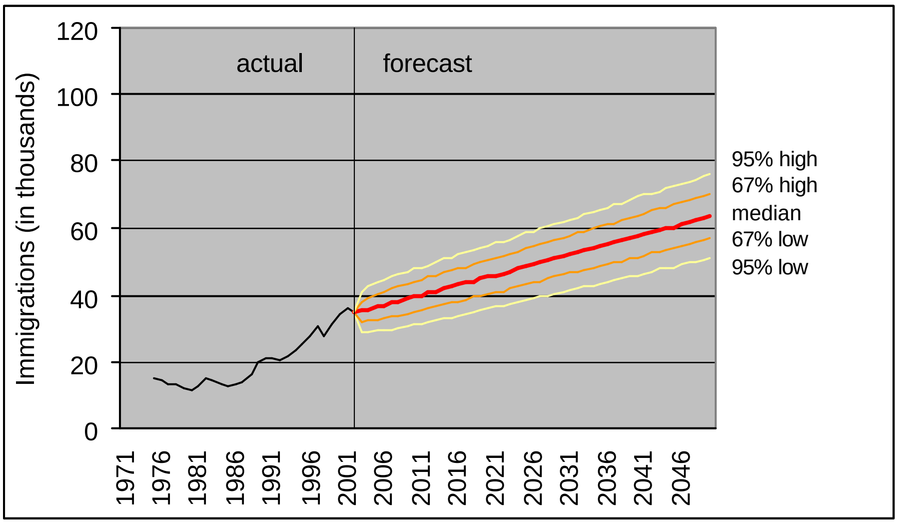


(b) Rest of Australia

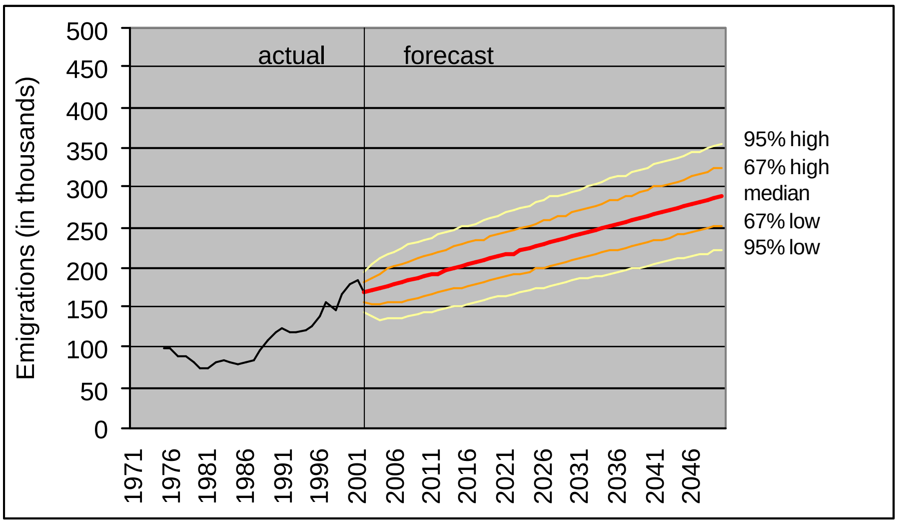


(c) Australia

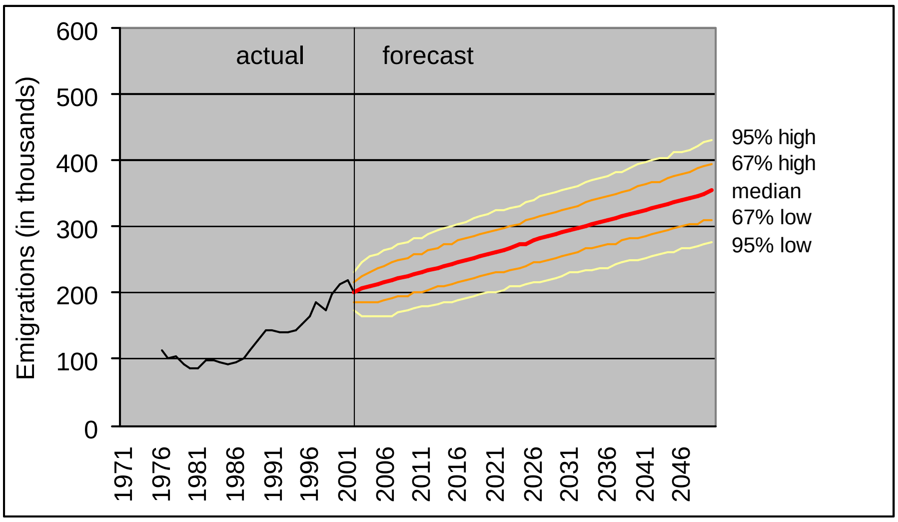
Figure 9 Observed and forecast immigration, 1971-2051



(a) Queensland

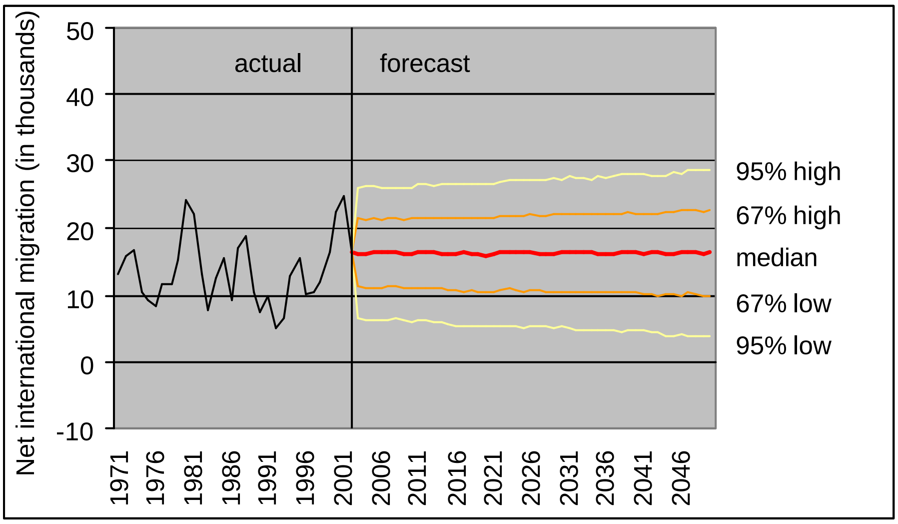


(b) Rest of Australia

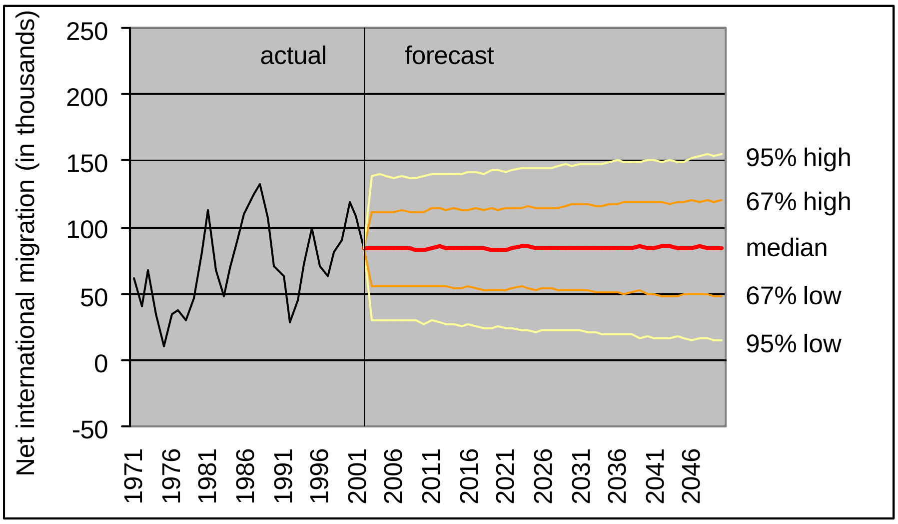


(c) Australia

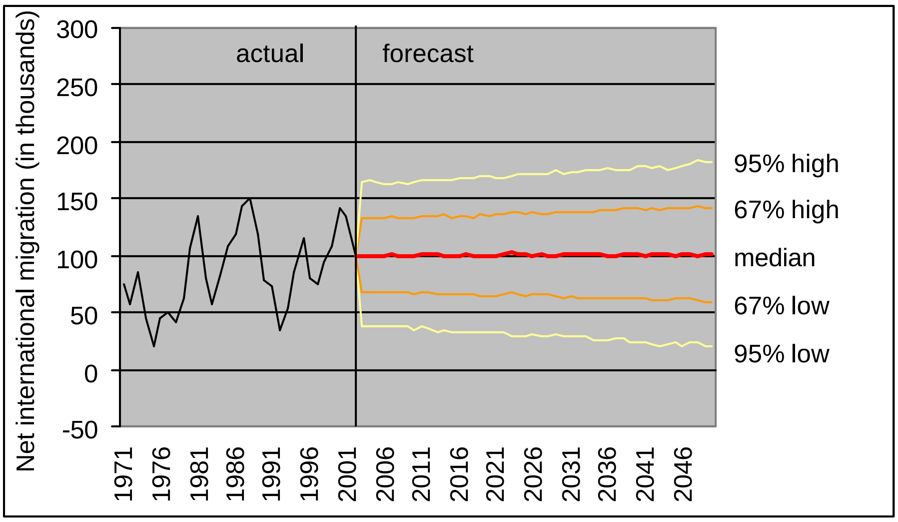
Figure 10 Observed and forecast emigration, 1971-2051



(a) Queensland



(b) Rest of Australia



(c) Australia

Figure 11 Observed and forecast net international migration